



UK Data Service

Annual report

April 2014 – March 2015



Data enriched research
Data enhanced impact



This second annual report of the UK Data Service highlights some of our key achievements. The case studies in particular demonstrate the breadth of the activities, but are only a fraction of what the service has to offer. It continues to provide access to high-quality data, to support to a wide range of users, and – as this Report clearly demonstrates – to make a major contribution to cutting-edge research and development programmes. Maintaining continuity while expanding into new business areas is a challenge for any organisation; the UK Data Service has risen successfully to this challenge.

Gordon Marshall
Chair, ESRC Data Infrastructure
Strategic Advisory Committee

The ESRC remains proud to fund the UK Data Service. The Director and his colleagues continue to maintain and develop a world-class service to researchers. The Service has been invaluable to the ESRC in advising on and promoting more open access to data and has been instrumental in the development of the ESRC's new Research Data Policy. It has also maintained a key role internationally, using its established expertise to support collaborative projects and offer advice to global partners. An Innovation Fund programme, bringing four exciting new resources into the core Service, has also demonstrated the ability of the Service to keep pace with a rapidly changing landscape, working with new partners in business and academia. That changing landscape has also seen the Service working increasingly with other ESRC investments, including the Big Data Network, making for a more efficient, coherent and robust data infrastructure.

Jane Elliott
Chief Executive, Economic and
Social Research Council

The constantly changing data landscape provides the UK Data Service with fascinating challenges. New and novel forms of data for social science research have the capacity to lead to new and novel insights into society, health, the economy and political behaviour. The creation, analysis and curation of these data also brings new problems—legal problems relating to both privacy and rights; technical problems relating to methods of analysis, storage, processing; methodological problems including the quality of the data, its relationship to the population and bias. The Service is continually grappling with these issues and not just these new and novel forms of data.

Matthew Woollard
Director, UK Data Service

Welcome to the second report of the activities of the UK Data Service

About us

The UK Data Service is funded by the Economic and Social Research Council (ESRC) to meet the data needs of academic researchers and students as well as data users and data owners in government, charities, foundations, think tanks and businesses. We hold a broad range of data for teaching and research including UK census data, large scale government surveys, longitudinal data, international macrodata, business microdata and qualitative studies.

We offer

Access to high quality regional, national and international social and economic data.

Support for policy-relevant research in the higher education, public and commercial sectors.

Development of best practice data preservation and sharing skills.

Information, training and support to improve data management skills.

Sharing of expertise with international data providers to remove barriers to accessing data.

Contents

Research 04

Enabling research that informs policy and improves practice

Skill-building 12

Teaching core data skills

Reach 14

Extending data access, support and expertise to new sectors and users

Infrastructure 22

Supporting future data technologies and research

Looking ahead 24

Preview and priorities for the coming year

Making access to high-quality, social and economic data and resources easier

The UK Data Service is increasing its collection of open data and continues to support data owners in making their data available as widely as possible.

Opening access to more data

Social science data pose a challenge in terms of open data as they often refer to living people. Our Data Access Policy protects the most sensitive social science data for research, while supporting data owners in making their data available as widely as possible. We view data access as a continuum, along which data collections may be moved to make them more open. Consent and anonymisation play a part at every stage of the data lifecycle, and protect the rights of the data subjects. We have 96 open data collections, which span international data, central government and large-scale survey data series, large-scale qualitative data collections and census data.

During this year we have renegotiated licences with data owning depositors to remove the need to register for access to 47 datasets including: Changes in the Structure of General Practice: the Patient's Viewpoint, 1977, Management of Productivity, 1984, Civil Rights in Schools: School Students' Views, 1997-1998 and the Prison Reading Survey, 1997. We also negotiated the removal of some of the more restrictive conditions of access that were attached to other collections, including the requirement that depositor permission be gained for access to another 32 datasets.

Most of the data available via our new Qualibank interface and our new ReShare online data self-deposit facility are open to all users without the need for registration. We use the

Open Government Licence or a relevant Creative Commons licence for open licensing and are making it easier for users to download open data using new functionality in our Discover interface.

We have also been undertaking the certification of some of our open data through the Open Data Institute, including the data that form part of our ESRC-funded Innovation Fund Programme, 'App Challenge', an exciting event which took place in summer 2015, which uses the European Quality of Life Time Series data.

International data

In January 2015, we were able to open up access to the International Monetary Fund (IMF) data which we distribute. These data had previously been restricted to staff and students at UK Further and Higher Education institutions, but thanks to a new policy introduced by the IMF whereby access to its online statistics databases became free to all users, we were able to remove all restrictions to accessing these data via the UKDS.Stat platform. These newly open data include the Government Finance Statistics, World Economic Outlook, International Financial Statistics, Direction of Trade Statistics and the Balance of Payments Statistics and are a great addition to the UK Data Service's open data collection.

Census data

For a prominent example of open data available from the UK Data Service, we need look no further than our census data. Here, all users can access 2011 and 2001 (UK) Census aggregate statistics under an Open Government Licence, as well as the UK Data Service back catalogue of aggregate data from the 1971, 1981 and 1991 Censuses. Most of the boundary data, available via census geography, are also open, as are most of the flow (origin-destination).

Opening up data to a wider audience is key to the Data Service vision, and we will maintain the same quality levels for these data.

Matthew Woollard, Director, UK Data Service

Providing a connected and accessible framework for 'big data'

The public and commercial sectors generate data at volume, speed and a level of complexity, which present challenges for their curation and reuse. These 'big data' offer a significant resource for research.

Big Data Network Support

The ESRC has funded three Big Data Network Research Centres:

- Urban Big Data Centre, University of Glasgow
- Business and Local Government Data Research Centre, University of Essex
- Consumer Data Research Centre, University of Leeds and University College London

Researchers need the right analytics framework to optimise the research potential of big data. The UK Data Service has been commissioned by the ESRC to support these Research Centres through the Service's Big Data Network Support (BDNS) programme, helping researchers to make the most of these data for knowledge exchange and impact.

Big data present challenges for existing models of data curation, preservation and access. We have developed methods to help identify those big data with relevance to socio-economic research with the aim of understanding, developing and promoting the opportunity of the research value of these complex datasets.

Our key activities include ensuring researchers have the skills they need to use these data. We are closing the skills gap by adding capacity to our user support and training team to help researchers in developing the additional skills they need to use these data. We are also investing in analytics, visualisation and insight training and offering specialist training courses for staff and researchers in these Research Centres.

The promotion of best practice in managing data for users of the Big Data Network is an important element of BDNS. Our focus includes helping data owners understand the benefits of making their data available, defining data purchase arrangements, supporting the Research Centres in gaining access to the data for research and defining arrangements for the preservation and reuse of these data. As part of this support, we have created an API that enables the Business and Local Government Data Centre to access the metadata held in the UK Data Service catalogue, so they can include a catalogue page in their own website so users do not need to be redirected to our Discover interface for accessing data and resources. This interface can easily be extended to each of the Big Data centres.

To provide a wide perspective on the big data landscape, especially for social media and transactional data, the Service is working with the Digital Preservation Coalition on two Technology Watch Reports which will be published in 2016.

Our involvement with the Big Data Network is resource intensive, but if it results in a more understandable framework for the curation and access to new and novel forms of data then there will be many more new research opportunities for using these data.

Helping people find, access, share and link the data we hold

Four innovative projects won funding through the UK Data Service's Innovation Fund. This fund was established by the ESRC as an incubator fund for new ideas and technology to be incorporated into the core Service.

UK Data Service Innovation Fund

As part of a 'Dragon's Den'-style event in July 2014, 15 competitors demonstrated how their idea had real potential for innovation in developing the Service and its resources. The UK Data Service is now implementing these projects with four successful applicants.

AudienceNet: Improving the presentation of Service data, resources and expertise

AudienceNet combine market research with connected technologies to provide insights. The AudienceNet team is helping us raise the profile of our data, resources and expertise by improving their presentation based on the findings of market research. We consulted current users of the Service and new and non-users from a range of sectors to develop insight into preferences for accessing our data and resources. We have been working on refreshing our website homepage, visualising our research and teaching case studies, an animation for new users, depositor videos and ways of visualising our data all of which will improve and simplify the user experience. These tools will help us to monitor trends in data usage to enhance our service.

Making research more accessible:

AudienceNet is also working with us to make the research carried out using data held by the UK Data Service more accessible. They have produced several infographic representations of research, such as this one on the safety of older adults crossing the road.

Crossing the road in time: finding the right pace for older adults

The ability to cross the road safely is important for the health of older people as without it they have less independence and less social interaction. The following case study examined the time it takes people aged 65+ to cross the road in time. The full case study is available at: <http://ukdataservice.ac.uk/use-data/data-in-use/case-study/?id=162>

1.2m/sec = 2.7mph
assumption for crossing times

informed by a small sample of healthy people crossing roads

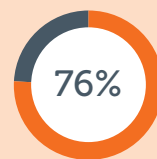
Among those able to walk 8' unaided and aged 65+



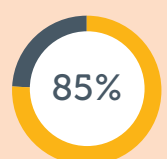
Less than 1.2 m/sec



Men



Women



Average Speed

0.9 m/sec

0.8 m/sec

Risk Utility Management software: Automating the removal of the risk of personal data disclosure

Data creators will soon have the ability to create and automate anonymised datasets from confidential sources through our Risk Utility Management software. The current method of creating anonymised datasets is a time-consuming and manual process, exerting significant costs on data producers who wish to generate anonymised versions of data. Developed in collaboration with our Norwegian sister service, NSD, this project is now undergoing testing with major survey owners and producers.

The easy-to-use interface enables the creation of anonymised datasets at the click of a button. A user uploads a source of confidential data and applies anonymisation preferences (based on an existing R tool, SDC Micro). Settings can be saved and reused, removing inconsistency. Other key features include a risk-utility score, allowing users to examine the effects of applying additional anonymisation techniques to the data. The Risk Utility Management project has sparked interest from a number of government departments in the UK, as well as the Government of Singapore which is keen to collaborate to produce a tool for use with its national data.

InFuseR: Advancing census data modelling and visualisation

Through collaboration with staff from the Centre for Spatial Demographics Research, at the University of Liverpool, our Census Support team has developed a public version of the InFuse API. This application will enable the creation of complex data queries spanning multiple tables, cells, geographies and censuses. It will offer iterative refinement of data access queries as a research project develops, enhanced sharing of research methods through sharing of data extraction code, research reproducibility, and enhanced data discovery through flexible filtering using the full metadata available in InFuse. It is expected that InFuse will contain ten billion counts relating to two million geographical areas.

Census data users can now use an open-source census data extraction and visualisation tool called InFuseR with a syntax-driven interface that interacts directly with our InFuse API. InFuseR is an R-based census aggregate data extractor, created by combining the open statistical software R with three resources housed by our Census Support team: census file metadata; detailed census data tables and the associated map boundaries. Access to 2011 Census aggregate statistics is free and open to users via InFuse or InFuseR, enabling them to build queries and extract data online. A redesigned InFuse home page has also made the application easier to use.



AppChallenge: Crowdsourcing exciting new apps using the Eurofund European Quality of Life Survey

Our open data crowdsourcing competition offers the opportunity for the development of innovative applications and services from developers who may not have otherwise discovered the UK Data Service. Offering a prize fund of £15,000, the #EULife @AppChallenge will crowdsource exciting new apps and ways of using the data provided from the Eurofound European Quality of Life Survey, which ran from 2003 to 2012.

The project involves a competition, for developers to create innovative uses for selected data made open via an API (Application Programming Interface), as certified by the Open Data Institute.

// First time this
important dataset is
being made available
via API/download //

*Emer Coleman, architect of
The London Datastore*

Using data intelligently: creating better knowledge for a better society

The Administrative Data Research Network provides a new gateway enabling researchers to access and link administrative data to form a more accurate and wider picture of our society.

Administrative Data Research Network

Administrative data are the information government departments and agencies collect when they register people and carry out transactions. The Administrative Data Service (ADS), set up in May 2014, coordinates the Network and serves as researchers' first point of contact. The ADS is a sister service to the UK Data Service and is hosted by the UK Data Archive. The ADS has developed a bespoke service for researchers who want access to link de-identified administrative data. To date, the Network's Approvals Panel has approved twenty projects and negotiations with the data owners are underway.

Incorporating existing expertise from the UK Data Archive the ADS has produced a series of policies and procedures that ensure a lawful, secure and efficient pathway for researchers to gain access to linked, de-identified administrative data collections. The Network's website <http://adrn.ac.uk> provides researchers with a 'one stop shop' where they can begin the process of using administrative data for their work. Two new web animations, targeted at the general public and research community, provide a clear explanation of the Network and of the security procedures for the Network.

SafePods

In partnership, the Administrative Data Research Centre Scotland and the Administrative Data Service have developed a national network of secure environments where researchers can gain access to administrative data. The first prototype "SafePod" has been installed in St. Andrews University and is currently undergoing user-testing.

Capacity building

ADS have developed national accreditation training for the Network in partnership with the UK Data Service, building on the Service's Secure Lab training. Over 150 researchers have been trained to date.

Network expertise

The Network consists of over 250 experts across the UK. The role of the ADS is to ensure Network consistency, promote collaboration, encourage efficiency, communicate consistently and to encourage integration across the wider data service landscape.



Administrative Data
Research Network

Bringing richly-described and contextualised data to users

Our QualiBank interface helps users to find and access qualitative data with an easy-to-use application.

QualiBank

Qualitative data provides insight into people's lives through public surveys, opinions, interviews and essays. The UK Data Service's QualiBank interface, launched this year, uses the latest technology to bring richly-described and contextualised data to users in a few quick and easy steps. Users can search and browse qualitative data in our collection by selecting any related object, for example, audio recordings or descriptions of photographs linked to the objects found, adding context and richness to the subject.

We are striving to make data citation commonplace and QualiBank allows a persistent citation to be made for selected extracts of data. Users can select an extract of text, based on one or more consecutive paragraphs and QualiBank dynamically retrieves a citation which can easily be pasted into research outputs.

Funded by the ESRC Digital Futures project, QualiBank now has 20 collections, all of which have been carefully chosen with advice from academic users and archivists, on the basis of their importance as research resources. Many of these data are open and include important collections such as:

- the British Archive of Political History, a project conducted between 1979 and 1980, which carried out interviews with key figures from politics, the civil service and the armed forces;
- Morale and Home Intelligence Reports, government documents containing rich descriptions of the attitudes and opinions of soldiers and the UK public during WWII;
- The Edwardians, a collection of over 450 oral histories of working class individuals from the Edwardian period; and
- the School Leavers Study, a collection of essays written by school children from the Isle of Sheppey in 1978, where they were asked to imagine that they were 60 years old and write reflective accounts of their lives.

While QualiBank serves as a successful demonstrator for access to qualitative data, it provides a unique citation facility which should advance the impact of data citation.

// Data citation is central to our understanding of the use of data in our collection, within and beyond academia, and where we can also begin to measure its impact by finding ways to derive evidence of reach and significance, as well as of relevance. **//**

*Victoria Moody, Director of Impact and Communications,
UK Data Service*

Supporting excellence in social science research

Data from our collection has been used in research that supports policy, stimulates debate and informs the public on key social science issues.

Data use in research, teaching and policy

We work with researchers to understand the innovative ways that our data are being used and develop case studies to communicate and promote the use of Service data in research, teaching and as a basis for informing policy. We publish these case studies on our website, and they can be searched through our Discover interface. These case studies demonstrate the impact of the whole of the Service. We are here to support excellent research, here are a few examples:

Energy use within the English and Welsh wine production industry

The English and Welsh wine production industry, with more than 120 wineries, has many challenges linked to its cool climate and emerging status as a quality wine-producing country. As a producer trying to establish itself in the competitive global wine market, the industry can benefit from a focus on energy sustainability as a key factor in reducing costs and achieving a sustainable business model. Mervyn Smyth from the University of Ulster, in collaboration with Alistair Nesbitt of Plumpton College, used data from the International Energy Agency World Energy Statistics to assess UK wine-making energy consumption against the industry average. They also surveyed a number of prominent vineyards in England and Wales to understand their energy usage.

Findings for policy:

The researchers found that a total of 512,350 kWh of energy was expended to produce a total of 1,032,194 bottles of wine in an average year from the surveyed wineries, demonstrating the potential for improved efficiency. The research has potential to support the adoption and application of benchmarks by winemaking associations, directly influencing the energy consumed in making wine in England and Wales to support sustainable winemaking.

Improved cancer survival rates amongst teenagers and young adults (TYA) in Northern England

Although cancer is relatively rare in teenagers and young adults, it is a major cause of death in this age group. Richard McNally and colleagues from Newcastle University investigated survival trends in teenagers and young adults diagnosed with cancer and living in Northern England between 1968 and 2008. The research used aggregate data from the UK 1971, 1981, 1991 and 2001 Censuses, along with data from the Northern Region Young Persons' Malignant Disease Registry.

Findings for policy:

The census datasets were key for assessing the demographic characteristics and deprivation measures of different areas in Northern England. The research was conducted in Northern England because it contains some of the most deprived areas of all the English regions and improvements in this region could reduce health inequalities. The research found that there have been large improvements in TYA cancer survival in Northern England over the last four decades.

Banking competition in Africa: sub-regional comparisons

The African banking sector has witnessed significant reforms over the last three decades following a long period of underperformance. Samuel Fosu from the University of Wolverhampton examined the extent of competition in African banking markets after several years of reforms and the extent to which competitive outcomes differ across sub-regional banking sectors. Drawing on data from the 2011 World Bank World Development Indicators, the research measured the degree of competition in African banking markets.

Findings for policy:

The research shows that sub-regional banking markets in Africa generally demonstrate monopolistic competitive behaviour, which is consistent with other emerging markets. These findings suggest that the recent financial sector reforms, such as liberalising interest rates and removing credit ceilings, have had significant impact on increasing competition. The findings provide a benchmark by which future changes in competition levels within the African banking sectors can be measured.

Does childhood mental health shape adult economic outcomes?

Evidence suggests that poor mental health in childhood has a pronounced and pervasive influence on people's chances of socioeconomic success as young adults. Understanding how childhood mental health shapes adult economic outcomes is a key question which affects health, educational and employment policies. Mark Egan and colleagues from the University of Stirling used data from two nationally-representative British cohort studies: the Longitudinal Study of Young People in England and the National Child Development Study to investigate the link between childhood psychological distress and higher youth unemployment.

The first study used the Longitudinal Study of Young People in England and looked at the level of psychological distress at the age of fourteen, as measured via self-report, using the General Health Questionnaire and the likelihood that the young person would be unemployed between the ages of sixteen and twenty-one. The results showed that those who were classified as highly distressed had unemployment rates two percentage points higher than their low distress peers.

In the second study, the research team used the National Child Development Study to test the robustness of the relationship between teacher-rated childhood distress at ages seven and eleven and greater difficulty entering the job market. This second study included control variable for the participants' childhood cognitive ability,

temperament and early childhood environment. Taken together, these two studies showed that children with high distress were 40% more likely to become unemployed. Using the National Child Development Study data the research team also found that children with high distress were 50% more likely to become unemployed after the onset of the 1980 recession in the UK.

Findings for policy:

This is the first study to conclusively demonstrate that highly distressed children experience youth unemployment levels higher than their less distressed peers. This association was not markedly affected by adjustment for adult distress, suggesting that childhood distress may set in motion a broad set of social, educational and health effects that act to influence later employment.

Further work on this issue

The research team suggest that future research should be carried out to examine whether the central finding of childhood distress robustly predicting unemployment replicates in other datasets and whether distressed young people are more likely to become unemployed during recessions. If this is the case, this suggests that there could be economic benefits to improving mental health services early in life.

This research was funded by the ESRC, Skills Development Scotland and the European Commission Marie Curie Initiative.

Our findings add to the growing literature which suggests that mental health, like cognitive abilities and socio-emotional skills, can be considered as an important factor involved in the production of economic success.

Mark Egan, Micheal Daly and Liam Delaney, Stirling University.

Developing best practice data preservation and sharing standards

We continue to lead the UK in supporting good data practice in research and optimising data sharing practices.

Supporting data creators through training

For data creators, we promote good data management practices, from the planning stages of research through to the deposit of data with a repository. Our guidance includes providing good annotation and documentation to make data independently understandable, and handling the ethical and legal challenges in managing and sharing data from research.

Our training programme ranges from short introductory webinars and face-to-face sessions, through to advanced two-day hands-on courses. A regular programme of data management events is held at the UK Data Service premises, complemented by bespoke training at conferences (ARMA conference, Blackpool, June 2014; Open Repositories Conference, Helsinki, June 2014).

This year we provided training at SOAS, Ghent University (Belgium) and for the Ecosystem Services and Poverty Alleviation programme during 2014. We also joined our US sister service ICPSR in delivering a 5-day summer school (July 2014) and delivered data management and deposit talks to researchers for the Nature Publishing Group and the London School of Hygiene and Tropical Medicine.

We have developed strong working relationships with a number of government data producers, including ONS, NatCen, the Home Office and key data producers for the ESRC, enabling us to promote good practice.

Our expertise in handling the data and metadata we receive and create ensures data are appropriately preserved and that they remain available, understandable and usable in the long term. We are committed to a transparent approach to ensure the continued trust of data producers and users. A part of this Trusted Digital Repositories (TDR) remit is met

by accreditation to the Data Seal of Approval (DSA) and continuing work to self-assess against ISO 16363 for Trusted Digital Repositories, which address governance, digital object management and infrastructure, and security risk management. This work aligns with and extends our existing information security management certification, ISO 27001.

ESRC Research Data Policy

The UK Data Service has worked closely with the ESRC to revise its Research Data Policy. The newly published ESRC Research Data Policy was released in March 2015, consisting of nine core principles which align with the RCUK Common Principles on Data Sharing. It also clearly outlines roles and responsibilities of all actors in the research data landscape. The key change to the Policy is that ESRC grant holders are no longer automatically required to offer data to the UK Data Service, though contracts with ESRC investments will explicitly state that the ESRC's data service provider must be used. Grant holders are still required to ensure that data created as a result of research projects are made available for reuse.

ESRC Framework for Research Ethics

The UK Data Service participated in the panel charged with revising and web-enabling the ESRC's Framework for Research Ethics. An updated version of the Framework for Research Ethics was released in January 2015 and encourages researchers to think ethically and emphasises the importance of identifying potential ethical issues throughout the research lifecycle of a project, including archiving and sharing of data.

Supporting data users through training and capacity building in data skills

Our expert staff train users in data access, use and management in a range of formats, including online webinars, face-to-face workshops and user conferences.

Supporting our users through training

Our training programme covers all aspects of the Service, including accessing data and skills for using the wide range of data we hold. We offer a rolling programme of core webinars to introduce different aspects of the Service and explain the key datasets we hold.

We collaborate with external experts to deliver training. One example of this is our webinar *The UK Data Service: An introduction to data on ageing*, where Dr Alan Marshall from the Frailty, Resilience and Inequality in Later Life project at the University of Manchester, highlights the use of the English Longitudinal Study of Ageing and the census.

We also work with data depositors such as the Office for National Statistics and NatCen to deliver at least five user conferences each year (for example the Health Surveys User Conference). These events prove extremely popular, providing an annual forum for data users and data depositors to discuss developments with key national surveys and census data.

Working with students and teachers to better understand their needs

The user support and training team have been working with students and teachers in higher education to gain a better understanding of their data and skills needs, and develop tailored on-line resources for this community. In addition to our routine annual stakeholder consultation exercise, we have conducted over thirty face-to-face and telephone interviews with undergraduates, final year postgraduates and higher education teachers covering topics including finding data, opinions on the current online resources and preferred formats for online resources.

This engagement has given us clear guidance on how best to meet their needs, one result being the development of a new suite of Student Resources webpages with essential information on *Getting Started with the UK Data Service*, FAQs, *Useful Data* and a set of short step-by-step guides to *Using Data for Your Dissertation*. We are also in the process of enhancing our *Teaching with Data* pages as a result of the consultation. Sharing beta versions of these pages with participants at a *Teaching With Data* workshop for Q-Step teachers, organised in collaboration with the Nuffield Foundation provided us with further insight into how participants use existing data in teaching settings. Topics discussed included the importance of real data, data sharing issues, software options at a time when R is becoming more popular, hands-on experience with Data Service resources and creating teaching materials and datasets.

“ I introduce students to the UK Data Service to show them the wide possibilities of available data that they can use for many different topics. As a teacher, the UK Data Service makes it easy to locate and use a wide range of data sources. I always tell my students that it is not necessary to go out and collect new data, but that we should re-use the fantastic data that other people have collected; this is the key value of the UK Data Service: readily available, high quality data. **”**

Anja Neundorf is Assistant Professor of Politics at the University of Nottingham

Working with researchers in the higher education, government and commercial sectors

Simplifying the depositing of data encourages data creators to share their data, increasing data resources available to researchers.

ReShare

Our new ReShare facility is an online self-deposit data repository for the archiving and sharing of research data, based on a customised version of EPrints, the open-source digital repository software. Data deposited in this way are typically those data that arise from ESRC research awards, the topics of which are spread widely across disciplines, covering both quantitative and qualitative data.

Researchers can find these data easily through the UK Data Service's Discover interface and can be confident that they meet with the necessary ethical and legal requirements, reviewed by the Service before their release.

We accompanied the launch of ReShare, with a webinar to showcase what this online repository has to offer. This webinar also provides a demonstration of the functionality of the Service and a practical process of depositing data into the repository. ReShare can be accessed via our website.

Investment in ReShare means we can really fulfil the new ESRC Research Data Policy.

[ReShare: first UK-based social science repository to be listed as a recommended repository by Scientific Data](#)

Scientific Data, the open-access data journal of Nature Publishing Group, publishes the Data Descriptor article type and recommends that datasets accompanying manuscripts be deposited in established and trusted repositories, such as the UK Data Service ReShare. The initiative ensures that these data are stably preserved for the longer-term, can be peer-reviewed, and will be easily accessible to the research community after publication.

The Service fully supports the concept of scientific transparency and is increasingly working with journals to help support their policies. The collaboration with Scientific Data gives social science and humanities researchers the opportunity to increase the discoverability of their data through submission of a Data Descriptor, whilst maintaining the Service's safeguarding of sensitive data.

Merging the HASSET and the ELSST: a single interface for content management

Two new interdependent web tools have been launched to manage the thesauri: SKOS-ELSTT and SKOS-HASSET.

Unifying resource discovery

Thesauri underpin resource discovery for the UK Data Service and CESSDA, the Consortium of European Social Science Data Archives. Although interdependent, the Humanities and Social Sciences Electronic Thesaurus (HASSET) and the European Language Social Science Thesaurus (ELSST) have different version releases and have always been maintained separately.

The innovative development of this single multilingual thesauri administration system ensures efficient content management of core terms. The system facilitates cooperation in thesaurus construction between CESSDA partners and the UK Data Service to ensure high quality, relevant keywords for cross-border data searches.

Widely and enthusiastically received throughout the international data community, the HASSET visualization application featured prominently in a recent International Society for Knowledge Organization (ISKO) debate concerning the future of the 'traditional' thesaurus. Leonard Will, a member of the working party for development of the British Standard 8723 for the Simple Knowledge Organisation System SKOS applied thesauri, suggested that the HASSET visualization tool was an example of the thesaurus of the future and the argument was taken up and heralded by other speakers at this event.

Stella Dextre Clarke, convenor of the working groups that developed BS 8723, reported on the debate in the ISKO UK by posting:

"To defeat any argument of the form "XXX has no place in YYY", all you have to provide is one counter-example. Just for starters: The UK Data Archive, powered by the HASSET thesaurus".

The increased interest in SKOS HASSET is reflected in greater demand for HASSET licences. From March 2011 to March 2013 before the application of SKOS to HASSET five licences were issued. From March 2013 to March 2015, after the application of SKOS, fourteen HASSET licences were issued.

These applications are innovative in the way in which they manage the relationships between connected thesaurus products, in relation to shared terms, equivalence relationships, the upload of suggestions for new concepts and the ways in which concepts may be moved between the two thesauri, with careful rules relating to structure and semantics built in.

HASSET and ELSST make our ever-expanding holdings more easily retrievable. They are the results of decades-worth of continued investment.

“ We are impressed with the quality of the thesaurus and also the visualisation and have not found anything comparable. ”

New South Wales Department of Family and Community Services

Sharing expertise with international data providers to remove barriers to access

Providing a framework for the deposit, preservation and distribution of data.

The UK Data Service now has concordats with three of the UK's major statistical organisations – the Office for National Statistics, National Records of Scotland and the Northern Ireland Statistical Research Agency. These concordats each provide single legal framework surrounding all areas of data custody – deposit, preservation and access to and distribution of data – owned by these organisations, and guidelines to maintain the quality of both the data and the level of service which we provide. Other data suppliers, including the Welsh Government, have expressed an interest in this concordat as a model agreement.

Enabling cross-national access to secure data

A reciprocal agreement with our partners at the Institute for Employment Research (IAB) in Germany has, for the first time, enabled cross-national access to secure data for UK and German researchers.

UK researchers can now access all the IAB's data from the UK Data Service Secure Lab while German researchers will be able to access Understanding Society data through a connection to the secure lab at the IAB. This agreement widens access to Understanding Society to researchers beyond the UK, but means the data never leave the country. These data have never been accessed from outside their respective countries owing to their confidential and sensitive nature, coupled with the unavailability of the secure technology required to provide robust access.

The agreement provides formal recognition by the UK Data Archive and IAB that their rules and procedures for access are equivalently rigorous – an important basis for negotiating transnational data access. Our agreement has the potential to be a model for use elsewhere, demonstrating our impact beyond the Service.

The project has been made possible thanks to the EU-funded project *Data without Boundaries*.

Away day 2015: identifying steps for success

February 2015 saw the latest UK Data Service whole staff away day.

“Bringing together our experts from across the service, the team examined successes from the last two years and identified priorities for expanding the reach and significance of the Service in the future. The away day also gave new and established staff the opportunity to network and learn from each other.”

Matthew Woollard, Director of the UK Data Service



Increasing the contribution data enriched research makes to the economy, society and quality of life

Supporting the development of research which addresses the issues that affect all our lives, both close to home and internationally.

Promoting our impact

Raising the visibility and use of the UK Data Service data and resources in teaching, research and by government, voluntary and community, and commercial sectors is a key focus of developing our impact.

Our priorities for impact are to:

- support the development of impact derived from research which uses Service data;
- engage more non-academic organisations and communities in using Service data, resources and expertise;
- articulate the role of the Service and its remit in terms of its contribution to wider societal benefit;
- expand innovation and collaboration in using Service data, resources and expertise between researchers and non-academic organisations and communities;
- formalise the demonstration and corroboration of the Service's impact; and
- communicate and promote the Service's impact

Communicating for impact

Attracting more people to our website to access our data, resources, expertise and to find out more about our impact is one element of enhancing communication through expanding our user base to new audiences, researchers and policymakers. Overall, the UK Data Service website is now attracting an average of 40,664 unique visitors each month, a 22% increase on the same period last year, with more users from many countries, over 6.5 million in the reporting period accessing our website, demonstrating our reach.

Our new Data Impact Blog, blog.ukdataservice.ac.uk was launched in January 2015. Contributions range from the research community, policy makers in the government and higher education sectors and UK Data Service staff. The blog is a resource for researchers, students, communities, policymakers, government and civil society and aims to encourage debate, share best practice and keep up to date with news, events and the latest data-driven impactful research and policymaking.

Understanding Society currently has 1,500 users. It's absolutely essential to have the UK Data Service ensure that such a wide range of people, mainly in the UK, but also internationally can use the data.

Michaela Benzeval, Deputy Director Understanding Society

Demonstrating our impact

We are improving how the Service formalises the demonstration of its impact, identifying opportunities to extend its reach and significance, and generating more and more authentic, user corroboration. Our focus is on enhancing the impact of research and teaching using Service data and resources, and of the Service as a whole through our infrastructure, expertise and resources. We will measure this impact through a focus on our benefit to skills, the economy, society, culture, public policy or services, health, the environment or quality of life, in terms of positive change.

Data in our collection have featured in a range of impact case studies in the Research Excellence Framework 2014, which allocates quality related funding to research active higher education institutions. These impact case studies demonstrate how data curated by the Service have supported research which has led to policy change in a range of areas including; the development of mathematical models for Practice based Commissioning budgets for adult mental health in the UK and informing public policy on obesity, both using the Health Survey for England. Service data have also informed the development of impact around understanding public attitudes towards the police and other legal institutions using the Crime Survey for England and Wales and research to support the development of the national minimum wage using the Labour Force Survey.

“ For us, it's not just about doing the research, it's about the impact of the research, so for NatCen, the UK Data Service is absolutely crucial, because as a charity, we pride ourselves on impact and a big part of that impact is getting the data out there. Without the UK Data Service there would be a hugely fragmented data landscape and fundamentally, many datasets wouldn't see the light of day and even datasets that were available would be hugely underused. **”**

Kirby Swales, Director of the Survey Centre at NatCen Social Research

Simplifying the navigation of census data

Making census data easier to navigate and manipulate, boosting the amount of data downloads.

1961 Census

As part of the development plan for the 1961 Census Aggregate Outputs Digital Recovery Project, the Census Support team has produced a report following an audit of the 130,000 scanned images of contemporary 1961 Census aggregate outputs publications supplied by ONS. There have been promising data recognition results from optical character recognition (OCR) of some initial samples of the set of scanned images of contemporary paper print-outs of 1961 Small Area Statistics provided by ONS.

Increasing census data downloads

The development of a number of easy-to-use tools, including the 'Easy Download' feature and 'look up' tables such as postcode directories, has made data navigation easier and has led to a huge increase in the number of downloads of census data sets. Most notably, there has been a 307% increase in downloads of Boundary Data via *Easy Download* compared with the previous 12 months. There has been an even greater increase in downloads of look-up tables.

Users can now also enjoy open access to the 2011 Census Flow Data (under Open Government Licence) via WICID, our flexible interface for all Origin-Destination Statistics (ODS). We have also been working closely with the Office for National Statistics to make available to users 205 safeguarded ODS tables and 22 open ODS tables from the 2011 Census.

Census user conference and historic microdata launch

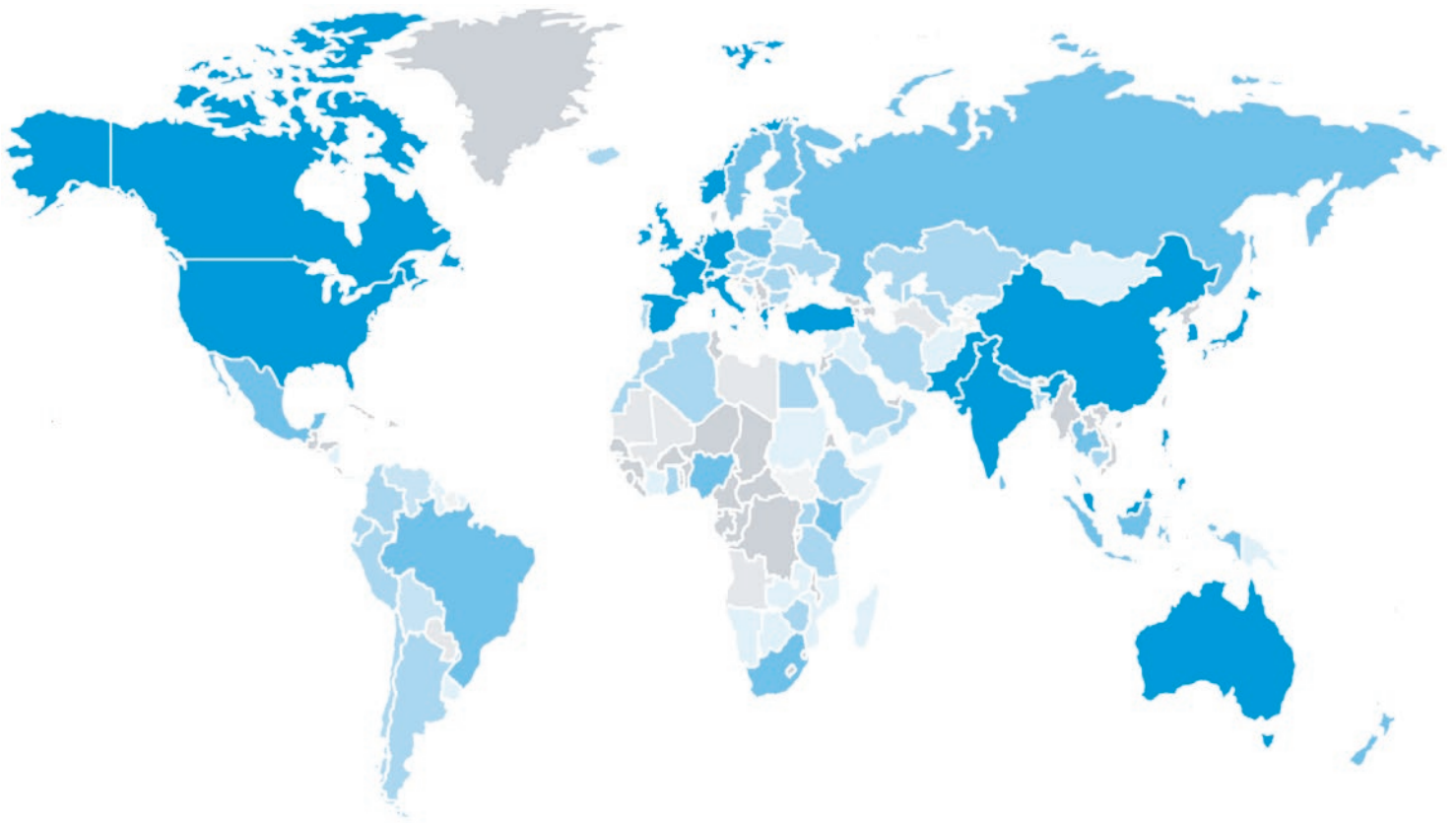
The Census Research User Conference, held at the Royal Statistical Society (RSS) in October 2014, was attended by 78 delegates who heard presentations by Census Support team experts and the UK census agencies on the latest developments in census data management and from users who have used these data in their research. The conference was followed by the Enhancing and Enriching Historic Census Microdata (EEHCM) event, which launched the new census microdata products from twentieth century censuses produced through the EEHCM project. This event marked the completion of the recovery of the first digital files containing census records and the production of microdata samples from the 1961, 1966, 1971 and 1981 censuses, through a collaboration between two of the UK Data Service host organisations, the UK Data Archive, and the Cathie Marsh Institute for Social Research.

UK Data Service
Census Support



Global usage of the UK Data Service

2014/15 saw an increase in the global use of the UK Data Service, with new users accessing our website from across the world. The USA and UK saw the largest number of new users, with more than 27,000 US-based and 260,000 UK-based individuals accessing the Service. This year also saw first time users from countries that have not previously used the Service, including Mali, Mauritania and Turkmenistan.



Developing and sharing our expertise in infrastructure

Working with partners across the data service community to provide the best experience for our service users.

CESSDA

The creation of a pan-European data service infrastructure has for a long time been considered a desirable catalyst for pan-European research. During this year work has been progressing on developing the new CESSDA (a consortium of European data services) and turning it into a fully-fledged European Research Infrastructure Consortium. The vision of CESSDA is to defragment the data services landscape of the European Research Area, integrating data, information and knowledge products and services by breaking down barriers which are brought about by language, culture and legal and institutional differences. A distributed infrastructure with harmonised standards and processes will provide social scientists with local access and support to this pan-European service. CESSDA will provide the closest solution to the free movement of social and economic data for scientific research across the whole of the ERA, and beyond, while respecting the key principles of subsidiarity and privacy. The UK Data Service is represented as a crucial partner in this project. Our Director, Matthew Woollard, is the Vice Chair of the CESSDA Board of Directors.

New international data platform: UKDS.Stat

Using OECD data warehousing technology, our new platform for international data, UKDS.Stat, was launched in January 2015. UKDS.Stat offers features such as animated time series charts, integrated metadata and the ability to save and share data queries (subsets of data). The international data portfolio is made up of over 60 databases, comprising nearly 450 different datasets and associated metadata. Established by the Organisation for Economic Co-operation and Development (OECD), the Statistical Information System Collaboration Community (SIS-CC) is a group of international institutions engaged in memoranda of understanding with the OECD for the purpose of sharing and co-developing statistical information systems, particularly the OECD's .Stat data dissemination platform.

The UK Data Service is delighted to be part of the SIS-CC and is working collaboratively with the OECD to develop UKDS.Stat. Other members of the group include: the Australian Bureau of Statistics, iStat (Italian Statistics), Statistics New Zealand, the European Commission, the National Bank of Belgium, Statistics Estonia and UNESCO. We have found collaborative development strengthens ties and relationships within the international statistical community and we have been able to benefit from each other's ideas and experiences. A project we are currently working on with the UNESCO Institute for Statistics (UIS) and the National Bank of Belgium (NBB) is to integrate Solr, an open source enterprise search platform, into .Stat to provide more advanced search features combined with significantly improved performance.

Underpinning the Development Data Revolution: Collaboration in Data Archiving - A joint ICSSR/UK Data Service/ESRC workshop

On 7-8 January 2015, the UK Data Service hosted a two-day knowledge exchange event with the Indian Council for Social Science (ICSSR) and the ESRC, bringing together key consultants involved in the planning of an Indian data service with UK Data Service experts. Participants shared their experiences and outlined how to establish and run a successful data service. The ICSSR is working with the Ministry of Statistics and Program Implementation and the Government of India to establish a Social Science Data Archive that will act as a single repository of social, economic and political data.

Such a service is reliant upon an archive that functions properly and the Open Archival Information System (OAIS) functional model shows just how curation activities are part of the glue that binds the Archive and the Service together. Core interactions with data producers and data users are key, including the workings of the pre-ingest function, ingest, archival storage, data management, preservation planning, administration, access and support. Yet to create a full data service, other aspects must also be taken into consideration, including acquisitions procedures, data appraisal, the required technical infrastructure, ethical and legal issues in data sharing and licensing implications, data storage options and appropriate back-up procedures.

The UK Data Service will be working with the ICSSR, assisting with the plans which are under development to come to fruition and allow researchers to have consolidated access to the great wealth of data which is created within the Indian Government.

ISO 27001 Certification

In March 2015 the UK Data Archive was subject to an external audit which resulted in recertification (for the second time) to ISO 27001 as well as transitioning from the 2005 to the 2013 version of the standard. The ISO 27000 family of standards helps organisations keep information assets secure. ISO 27001:2013 places a much greater emphasis on security within the software development process, with the introduction of new controls. Maintaining this certification is crucial to our ability to continue to provide access to data which has a disclosure risk.



Looking ahead

This report demonstrates the wide range of projects we have developed to improve the experience of UK Data Service for users and depositors and to attract new users from a range of sectors.

There is more to come – the UK Data Service’s team of experts has already begun work on some new and exciting initiatives, on which we can report outputs and impact in next year’s Annual Report. Take a look at some of the work we have underway:

ESRC Human Rights workshop

We look forward to the outcome of our ESRC Human Rights data workshop, which is aimed at those working in civil society organisations that collect administrative, monitoring or evaluation data and who are charged with overseeing strategy and campaigns. To be held in October 2015, the workshop will help define the ESRC’s next phase of civil society engagement. It will consider how civil society organisations can link up and exploit academic practices to create outcomes that make a difference, what practical toolkits might be developed and what the most useful models for capacity building should be.

App Challenge

We are keenly awaiting the outcome of the summer 2015 @AppChallenge, to crowdsource exciting new apps and ways of using the data provided from the Eurofound European Quality of Life Survey, which ran from 2003 to 2012 to see what innovative uses and outlets for data are put forward by the developers taking part.

Open version of GeoConvert

An open version of GeoConvert has been launched using open postcode estimates from the 2011 UK census together with digital boundary data from the 1991, 2001 and 2011 census, all of which are available under an Open Government Licence. The new open version enables the conversion between 1991, 2001 and 2011 census geographies as well as the ability to link postcode data with the latest deprivation measures for England, Scotland, Wales and Northern Ireland and the 2011 output area classification. We have also added the coordinate data for postcodes for download. The coordinates are available in British national grid format for England, Wales and Scotland, Irish Transverse Mercator for Northern Ireland and also latitude and longitude to enable mapping in various online applications such as Google Maps.

CLOSER Portal

Cohort and Longitudinal Studies Enhancement Resources (CLOSER) funded by the ESRC and the Medical Research Council (MRC), aims to maximise the use, value and impact of the UK's longitudinal studies internationally. The project works to stimulate interdisciplinary research, develop shared resources, provide training and share expertise. We are working with the Centre for Longitudinal Studies and other major study owners to develop a Portal where users can search across the cohort and panel studies to find variables and questions.

Forthcoming guides and workshops to good practice

Over the next year the Service will publish three good practice guides. The first will be on 'Sharing social data in multidisciplinary, multi-stakeholder research', which will be a sister guide to our popular 'Depositing Shareable Survey Data' guide published in 2014. The second will be aimed at new data depositors and will outline 'What we do for you', with the aim of sharing the Service's facilities, expertise and support for depositors. The last one will be on the Service's approach to principles that support best practices surrounding access to confidential microdata, based around the '5 Safes' model, and will set out the Service's approach to disclosure review of microdata.

We will also be running a major workshop in September on setting up secure research facilities, entitled 'The 5 Safes of secure access to confidential data'.

FOSTER Training

We are working with partners of the Consortium of European Social Science Data Archives (CESSDA) to deliver a series of hands-on multi-day doctoral training events for the social sciences in Research Data Management for Open Data. The Facilitate Open Science Training For European Research (FOSTER) training project will be held at universities in five European countries, to set in place sustainable mechanisms for EU researchers to foster open science. Training will cover critical areas related to research data management for open data in the social sciences, such as metadata and contextual description, ethical and legal aspects of sharing sensitive or confidential data, anonymising research data for reuse, and writing a data management plan.

International data

We are planning a major upgrade to UKDS.Stat (v.7) to include an API which will provide access to an increasing number of datasets. This will enable developers to call the API programmatically, permitting data to be integrated into programs and web sites. The API uses the SDMX-JSON format; an international format for statistical data and metadata designed to enable web developers to easily integrate data into web sites or web applications.

Testing and integrating big data activities

Work will continue to ensure that the ESRC's investment in the Big Data Network is fully supported. We will be promoting best practices that will add value to the network, working with the Digital Preservation Coalition to gain a wider perspective on the big data landscape especially as it relates to curation and reuse. We will be integrating our secure training activities across the network, and investigating how to repurpose some of our established data service activities so that they best meet the needs of the users who use these data in their research.

Who's who

The UK Data Service is delivered by staff with specialist expertise in research data based at higher education institutions across the UK.

UK Data Archive, University of Essex

Lead organisation of the UK Data Service and an internationally acknowledged centre of expertise in acquiring, curating and providing access to data. The Archive has been providing coordinated data service infrastructure to the ESRC for almost 50 years. It also hosts the Administrative Data Service, and carries out research and development in other areas of data service infrastructure.

Jisc

Jisc is a registered charity which champions the use of digital technologies in UK education and research. Jisc hosts a significant number of the UK's research information assets and builds applications to help people make the most of these rich resources. Through the UK Data Service Jisc provides access to, and specialist support for, the databanks of such prestigious organisations as the International Monetary Fund, the World Bank and OECD as well as aggregate statistics from the 1971 to 2011 UK censuses.

Cathie Marsh Institute for Social Research (CMIST), University of Manchester

A centre of research excellence specialising in the application of advanced quantitative methods in an interdisciplinary social science context. CMIST plays a key role in the support and development of UK secondary microdata including the Sample of Anonymised Records (SARs) from the census.

School of Geography, University of Leeds

The School has a long-standing association with both the spatial analysis of census data and the development of web-based systems to ensure their use in the wider academic community.

Geography and the Environment, University of Southampton

The Population, Health and Wellbeing research group has an outstanding reputation for leadership in population and health research, based on the combination of innovation in methodology covering geographical information systems (GIS), spatial analysis and quantitative and qualitative methods with substantive expertise in census, population and health issues. Their work comprises spatial population analysis and modelling; cultures, spaces and practices of care and population health.

EDINA, University of Edinburgh

EDINA acts as the ESRC's specialist geography unit for the UK census, supporting UK Data Service users in accessing and using the geography outputs for the census.

Department of Information Studies, University College London

An international centre for knowledge creation and transfer in the fields of librarianship, archives and records management, publishing, information science and digital humanities.

Centre for Advanced Spatial Analysis, University College London

One of the leading forces in the science of cities, generating new knowledge and insights for use in city planning, policy and design and drawing on the latest geospatial methods and ideas in computer-based visualisation and modelling.

Governing Board

In 2014 the Governing Board was replaced by the ESRC Data Infrastructure Strategic Advisory Committee (DISAC).

DISAC

Chair

Gordon Marshall
Leverhulme Trust

Vice-Chair

Simon Briscoe
independent

Members

Fiona Armstrong
ESRC

Roger Burrows
Goldsmiths, University of London

Gerald Donnelly
National Records of Scotland

Guy Goodwin
Office for National Statistics

David Marshall
Northern Ireland Statistics and Research Agency

Maurizio Pilu
Connected Digital Economy Catapult

Gesche Schmid
Local Government Association

Sharon Witherspoon
Nuffield Foundation

Governing Board

(until November 2014)

Chair

Gordon Marshall
Leverhulme Trust

Vice-Chair

Simon Briscoe
independent

Members

George Alter
ICPSR, University of Michigan

Fiona Armstrong
ESRC

Keith Brown
University of Manchester

Rachel Bruce
Jisc

John Curtice
University of Strathclyde

Keith Dugmore
Demographic Decisions

Peter Elias
University of Warwick

Nigel Fielding
University of Surrey

Guy Goodwin
Office for National Statistics

Todd Landman
University of Essex

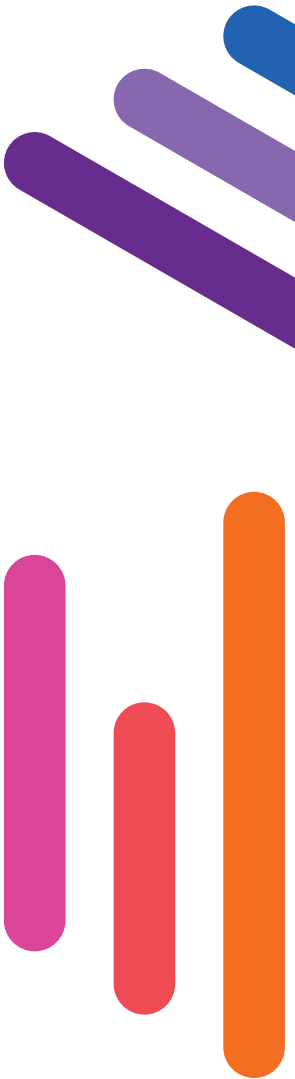
Paul Longley
University College London

David Marshall
Northern Ireland Statistics and Research Agency

Andrew Morris
University of Dundee

Sharon Witherspoon
Nuffield Foundation

ukdataservice.ac.uk



UK Data Service



Funding provided by



Contains recycled materials